Derivation of Kalman Filtering and Smoothing Equations

Byron M. Yu
Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA
byronyu@stanford.edu

Krishna V. Shenoy
Department of Electrical Engineering
Neurosciences Program
Stanford University
Stanford, CA 94305, USA
shenoy@stanford.edu

Maneesh Sahani
Gatsby Computational Neuroscience Unit
University College London
17 Queen Square, London WC1N 3AR, UK
maneesh@gatsby.ucl.ac.uk

October 19, 2004

Abstract

The Kalman filtering and smoothing problems can be solved by a series of forward and backward recursions, as presented in [1]–[3]. Here, we show how to derive these relationships from first principles.

1 Introduction

We consider linear time-invariant dynamical systems (LDS) of the following form:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_t \tag{1}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t \tag{2}$$

where $\mathbf{x}_t$ and $\mathbf{y}_t$ are the state and output, respectively, at time t. The noise terms, $\mathbf{w}_t$ and $\mathbf{v}_t$, are zero-mean normally-distributed random variables with covariance matrices Q and R, respectively. The initial state, $\mathbf{x}_1$, is normally-distributed with mean $\boldsymbol{\pi}_1$ and variance V_1.

In this work, we assume that the parameters of the linear dynamical system, namely A, C, Q, R, $\boldsymbol{\pi}_1$, and V_1 are known. Whereas the outputs are observed, the state and noise variables are hidden.

The goal is to determine $P(\mathbf{x}_t|\{\mathbf{y}\}_1^t)$ and $P(\mathbf{x}_t|\{\mathbf{y}\}_1^T)$ for $t = 1, \ldots, T$. These are the solutions to the filtering and smoothing problems, respectively. Both distributions are normally-distributed for the system described by (1) and (2), so it suffices to find the mean and variance of each distribution.

We will use the same notation as in [1]. $E(\mathbf{x}_t|\{\mathbf{y}\}_1^\tau)$ is denoted by $\mathbf{x}_t^\tau$ and $\text{Var}(\mathbf{x}_t|\{\mathbf{y}\}_1^\tau)$ is denoted by V_t^τ. The sequence of T outputs $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$ is denoted by $\{\mathbf{y}\}$. A subsequence of outputs $(\mathbf{y}_{t_0}, \mathbf{y}_{t_0+1}, \ldots, \mathbf{y}_{t_1})$ is denoted by $\{\mathbf{y}\}_{t_0}^{t_1}$.

2 Forward Recursions: Filtering

By the assumptions of the LDS described by (1) and (2), $P(\mathbf{x}_t|\{\mathbf{y}\}_1^t)$ is a normal distribution. We seek its mean $\mathbf{x}_t^t$ and variance V_t^t.

$$\begin{aligned}
\log\, P(\mathbf{x}_t|\{\mathbf{y}\}_1^t) &= \log\, P(\mathbf{x}_t|\{\mathbf{y}\}_1^{t-1}, \mathbf{y}_t) \\
&= \log\, P(\mathbf{y}_t|\mathbf{x}_t, \{\mathbf{y}\}_1^{t-1}) + \log\, P(\mathbf{x}_t|\{\mathbf{y}\}_1^{t-1}) + \ldots \\
&= \log\, P(\mathbf{y}_t|\mathbf{x}_t) + \log\, P(\mathbf{x}_t|\{\mathbf{y}\}_1^{t-1}) + \ldots \\
&= -\frac{1}{2}(\mathbf{y}_t - C\mathbf{x}_t)' R^{-1}(\mathbf{y}_t - C\mathbf{x}_t) - \frac{1}{2}(\mathbf{x}_t - \mathbf{x}_t^{t-1})'(V_t^{t-1})^{-1}(\mathbf{x}_t - \mathbf{x}_t^{t-1}) + \ldots \\
&= -\frac{1}{2}\mathbf{x}_t' \left(C'R^{-1}C + (V_t^{t-1})^{-1}\right)\mathbf{x}_t + \mathbf{x}_t'\left(C'R^{-1}\mathbf{y}_t + (V_t^{t-1})^{-1}\mathbf{x}_t^{t-1}\right) + \ldots \qquad (3)
\end{aligned}$$

Note that, in general, if $\mathbf{z}$ is normally-distributed with mean $\boldsymbol{\mu}$ and variance Σ,

$$\begin{aligned}
\log\, P(\mathbf{z}) &= -\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu}) + \ldots \\
&= -\frac{1}{2}\mathbf{z}'\Sigma^{-1}\mathbf{z} + \mathbf{z}'(\Sigma^{-1}\boldsymbol{\mu}) + \ldots \qquad (4)
\end{aligned}$$

Comparing the first terms in (3) and (4) and using the Matrix Inversion Lemma,

$$\begin{aligned}
V_t^t &= \left(C'R^{-1}C + (V_t^{t-1})^{-1}\right)^{-1} \\
&= V_t^{t-1} - K_t C V_t^{t-1} \qquad (5)
\end{aligned}$$

where

$$K_t = V_t^{t-1}C'\left(R + CV_t^{t-1}C'\right)^{-1}. \qquad (6)$$

To find the time update for the variance, we use the fact that $A\mathbf{x}_{t-1}$ and $\mathbf{w}_{t-1}$ are independent

$$\begin{aligned}
V_t^{t-1} &= \text{Var}(A\mathbf{x}_{t-1}|\{\mathbf{y}\}_1^{t-1}) + \text{Var}(\mathbf{w}_{t-1}|\{\mathbf{y}\}_1^{t-1}) \\
&= AV_{t-1}^{t-1}A' + Q. \qquad (7)
\end{aligned}$$

Before finding the mean of the normal distribution, we derive the following matrix identity

$$\begin{aligned}
(A+B)^{-1}(A+B) &= I \\
I - (A+B)^{-1}A &= (A+B)^{-1}B \\
(I - (A+B)^{-1}A)B^{-1} &= (A+B)^{-1}. \qquad (8)
\end{aligned}$$

Comparing the second terms in (3) and (4) and applying the matrix identity (8),

$$\begin{aligned}
\mathbf{x}_t^t &= V_t^t\left(C'R^{-1}\mathbf{y}_t + (V_t^{t-1})^{-1}\mathbf{x}_t^{t-1}\right) \\
&= V_t^{t-1}C'\left(I - \left(R + CV_t^{t-1}C'\right)^{-1}CV_t^{t-1}C'\right)R^{-1}\mathbf{y}_t + (I - K_tC)\mathbf{x}_t^{t-1} \\
&= V_t^{t-1}C'\left(R + CV_t^{t-1}C'\right)^{-1}\mathbf{y}_t + (I - K_tC)\mathbf{x}_t^{t-1} \\
&= K_t\mathbf{y}_t + (I - K_tC)\mathbf{x}_t^{t-1} \\
&= \mathbf{x}_t^{t-1} + K_t(\mathbf{y}_t - C\mathbf{x}_t^{t-1}). \qquad (9)
\end{aligned}$$

The time update for the mean can be found by conditioning on $\mathbf{x}_{t-1}$

$$\begin{aligned}
\mathbf{x}_t^{t-1} &= E_{\mathbf{x}_{t-1}}\left(E(\mathbf{x}_t|\mathbf{x}_{t-1},\{\mathbf{y}\}_1^{t-1})\,\middle|\{\mathbf{y}\}_1^{t-1}\right) \\
&= E_{\mathbf{x}_{t-1}}\left(A\mathbf{x}_{t-1}\,\middle|\{\mathbf{y}\}_1^{t-1}\right) \\
&= A\mathbf{x}_{t-1}^{t-1}. \qquad (10)
\end{aligned}$$

The recursions start with $\mathbf{x}_1^0 = \boldsymbol{\pi}_1$ and $V_1^0 = V_1$. Equations (5), (6), (7), (9), and (10) together form the Kalman filter forward recursions, as shown in [1].

3 Backward Recursions: Smoothing

Like the filtered posterior distribution $P(\mathbf{x}_t|\{\mathbf{y}\}_1^t)$, the smoothed posterior distribution $P(\mathbf{x}_t|\{\mathbf{y}\}_1^T)$ is also normal. We seek its mean $\mathbf{x}_t^T$ and variance V_t^T. We are also interested in the covariance of the joint posterior distribution $P(\mathbf{x}_{t+1},\mathbf{x}_t|\{\mathbf{y}\}_1^T)$, denoted $V_{t+1,t}^T$.

$$\begin{aligned}
\log\, P(\mathbf{x}_{t+1},\mathbf{x}_t|\{\mathbf{y}\}_1^T) &= \log\, P(\mathbf{x}_t|\mathbf{x}_{t+1},\{\mathbf{y}\}_1^T) + \log\, P(\mathbf{x}_{t+1}|\{\mathbf{y}\}_1^T) \\
&= \log\, P(\mathbf{x}_t|\mathbf{x}_{t+1},\{\mathbf{y}\}_1^t) + \log\, P(\mathbf{x}_{t+1}|\{\mathbf{y}\}_1^T) \\
&= \log\, P(\mathbf{x}_{t+1}|\mathbf{x}_t) + \log\, P(\mathbf{x}_t|\{\mathbf{y}\}_1^t) - \log\, P(\mathbf{x}_{t+1}|\{\mathbf{y}\}_1^t) + \log\, P(\mathbf{x}_{t+1}|\{\mathbf{y}\}_1^T) \\
&= -\frac{1}{2}(\mathbf{x}_{t+1}-A\mathbf{x}_t)'Q^{-1}(\mathbf{x}_{t+1}-A\mathbf{x}_t) - \frac{1}{2}(\mathbf{x}_t-\mathbf{x}_t^t)'(V_t^t)^{-1}(\mathbf{x}_t-\mathbf{x}_t^t) \\
&\quad + \frac{1}{2}(\mathbf{x}_{t+1}-\mathbf{x}_{t+1}^t)'(V_{t+1}^t)^{-1}(\mathbf{x}_{t+1}-\mathbf{x}_{t+1}^t) \\
&\quad - \frac{1}{2}(\mathbf{x}_{t+1}-\mathbf{x}_{t+1}^T)'(V_{t+1}^T)^{-1}(\mathbf{x}_{t+1}-\mathbf{x}_{t+1}^T) + \ldots \\
&= -\frac{1}{2}\mathbf{x}_{t+1}'\left(Q^{-1}-(V_{t+1}^t)^{-1}+(V_{t+1}^T)^{-1}\right)\mathbf{x}_{t+1} \\
&\quad - \frac{1}{2}\mathbf{x}_{t+1}'(-Q^{-1}A)\mathbf{x}_t - \frac{1}{2}\mathbf{x}_t'(-A'Q^{-1})\mathbf{x}_{t+1} \\
&\quad - \frac{1}{2}\mathbf{x}_t'\left(A'Q^{-1}A+(V_t^t)^{-1}\right)\mathbf{x}_t + \mathbf{x}_t'\left((V_t^t)^{-1}\mathbf{x}_t^t\right) + \ldots \qquad (11)
\end{aligned}$$

Note that, in general, if $[\mathbf{z}_1' \;\; \mathbf{z}_2']'$ is normally-distributed with mean $[\boldsymbol{\mu}_1' \;\; \boldsymbol{\mu}_2']'$, then the log density can be expressed in the form

$$\begin{aligned}
\log\, P(\mathbf{z}_1,\mathbf{z}_2) &= -\frac{1}{2}\begin{bmatrix}\mathbf{z}_1-\boldsymbol{\mu}_1 \\ \mathbf{z}_2-\boldsymbol{\mu}_2\end{bmatrix}'\begin{bmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}\begin{bmatrix}\mathbf{z}_1-\boldsymbol{\mu}_1 \\ \mathbf{z}_2-\boldsymbol{\mu}_2\end{bmatrix} + \ldots \\
&= -\frac{1}{2}\mathbf{z}_1'S_{11}\mathbf{z}_1 - \frac{1}{2}\mathbf{z}_1'S_{12}\mathbf{z}_2 - \frac{1}{2}\mathbf{z}_2'S_{21}\mathbf{z}_1 - \frac{1}{2}\mathbf{z}_2'S_{22}\mathbf{z}_2 + \mathbf{z}_2'\left(S_{21}\boldsymbol{\mu}_1+S_{22}\boldsymbol{\mu}_2\right) + \ldots \qquad (12)
\end{aligned}$$

The covariance of $[\mathbf{z}_1' \;\; \mathbf{z}_2']'$ is

$$\begin{aligned}
\begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} &= \begin{bmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}^{-1} \\
&= \begin{bmatrix}F_{11}^{-1} & -F_{11}^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}F_{11}^{-1} & F_{22}^{-1}\end{bmatrix} \qquad (13) \\
&= \begin{bmatrix}S_{11}^{-1}+S_{11}^{-1}S_{12}F_{22}^{-1}S_{21}S_{11}^{-1} & -F_{11}^{-1}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}F_{11}^{-1} & S_{22}^{-1}+S_{22}^{-1}S_{21}F_{11}^{-1}S_{12}S_{22}^{-1}\end{bmatrix}, \qquad (14)
\end{aligned}$$

where

$$F_{11} = S_{11} - S_{12}S_{22}^{-1}S_{21}$$
$$F_{22} = S_{22} - S_{21}S_{11}^{-1}S_{12}.$$

Comparing the first four terms in (11) and (12), we can write

$$\begin{bmatrix} V_{t+1}^T & V_{t+1,t}^T \\ V_{t,t+1}^T & V_t^T \end{bmatrix} = \begin{bmatrix} Q^{-1} - (V_{t+1}^t)^{-1} + (V_{t+1}^T)^{-1} & -Q^{-1}A \\ -A'Q^{-1} & A'Q^{-1}A + (V_t^t)^{-1} \end{bmatrix}^{-1} \tag{15}$$

We first simplify two expressions that will appear when inverting the block matrix in (15). First, using the Matrix Inversion Lemma,

$$\begin{aligned} S_{22}^{-1} &= \left(A'Q^{-1}A + (V_t^t)^{-1}\right)^{-1} \\ &= V_t^t - V_t^t A'(V_{t+1}^t)^{-1}AV_t^t \\ &= V_t^t - J_t V_{t+1}^t J_t', \end{aligned} \tag{16}$$

where we define

$$J_t = V_t^t A'(V_{t+1}^t)^{-1}. \tag{17}$$

Second, applying the matrix identity (8),

$$\begin{aligned} S_{22}^{-1}S_{21} &= -(V_t^t - J_t V_{t+1}^t J_t')A'Q^{-1} \\ &= -V_t^t A'\left(I - (Q + AV_t^t A')^{-1}AV_t^t A'\right)Q^{-1} \\ &= -V_t^t A'(Q + AV_t^t A')^{-1} \\ &= -J_t. \end{aligned} \tag{18}$$

Now, we invert the block matrix in (15). Using (16),(18), and the fact that $F_{11}^{-1} = V_{t+1}^T$ from (13),

$$\begin{aligned} V_t^T &= S_{22}^{-1} + S_{22}^{-1}S_{21}F_{11}^{-1}S_{12}S_{22}^{-1} \\ &= (V_t^t - J_t V_{t+1}^t J_t') + (-J_t)V_{t+1}^T(-J_t') \\ &= V_t^t + J_t(V_{t+1}^T - V_{t+1}^t)J_t' \end{aligned} \tag{19}$$

and

$$\begin{aligned} V_{t+1,t}^T &= -F_{11}^{-1}S_{12}S_{22}^{-1} \\ &= V_{t+1}^T J_t'. \end{aligned} \tag{20}$$

Using (17), (19), and (20), we can also derive a recursive formulation for the covariance

$$\begin{aligned} V_{t,t-1}^T &= V_t^T J_{t-1}' \\ &= \left(V_t^t + J_t(V_{t+1}^T - V_{t+1}^t)J_t'\right)J_{t-1}' \\ &= \left(V_t^t + J_t(V_{t+1,t}^T - AV_t^t)\right)J_{t-1}' \\ &= V_t^t J_{t-1}' + J_t(V_{t+1,t}^T - AV_t^t)J_{t-1}'. \end{aligned} \tag{21}$$

Using (5), (17), and (20), this recursion is initialized with

$$\begin{aligned} V_{T,T-1}^T &= V_T^T J_{T-1}' \\ &= (I - K_T C)V_T^{T-1}J_{T-1}' \\ &= (I - K_T C)AV_{T-1}^{T-1}. \end{aligned} \tag{22}$$

To find the mean, we compare the last terms in (11) and (12). Using (16), (17), and (18),

$$\begin{aligned} S_{21}\mathbf{x}_{t+1}^T + S_{22}\mathbf{x}_t^T &= (V_t^t)^{-1}\mathbf{x}_t^t \\ \mathbf{x}_t^T &= -S_{22}^{-1}S_{21}\mathbf{x}_{t+1}^T + S_{22}^{-1}(V_t^t)^{-1}\mathbf{x}_t^t \\ &= J_t\mathbf{x}_{t+1}^T + (I - J_tA)\mathbf{x}_t^t \\ &= \mathbf{x}_t^t + J_t(\mathbf{x}_{t+1}^T - A\mathbf{x}_t^t). \end{aligned} \tag{23}$$

Equations (17), (19), (21), (22), and (23) together form the Kalman smoother backward recursions, as shown in [1]. Equivalently, (20) can be used in the place of (21) and (22) to reduce the computation required to find the covariance.

References

[1] Z. Ghahramani and G.E. Hinton. Parameter estimation for linear dynamical systems. Technical Report CRG-TR-96-2, University of Toronto, 1996.

[2] R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264, 1982.

[3] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.